Genome Wide Identification, Classification and Functional Characterization of IQD Gene Family in Diploid and Tetraploid Cotton (Gossypium)

Abdul Rehman

Zhengzhou University

Zhen Peng

Institute of Cotton Research Chinese Academy of Agricultural Science

Hongge Li

Institute of Cotton Research Chinese Academy of Agricultural Science,

Guangyong Qin

Zhengzhou University

Yinhua Jia

Institute of Cotton Research Chinese Academy of Agricultural Science

Zhaoh Pan

Institute of Cotton Research Chinese Academy of Agricultural Science

Shoupu He

Zhengzhou University

Abdul Qayyum

Bahauddin Zakariya university

Xiongming Du (✉ cottonchina8848@126.com)

Zhengzhou University

Manuscript

Keywords: Circos, Synteny, Collinearity, Cotton, Expression, Orthologous genes

Posted Date: February 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-166037/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Aims

Calmodulin (CaM) is considered as the most significant Ca^{2+} signaling messenger that mediate various biochemical and physiological reactions. Whereas the calcium level can be provoked by extracellular or intracellular stimuli. CaM is activated by calcium binding proteins (CaMBPs) to regulate physiology of the cell. IQ domain (IQD) proteins are plant specific CML/CaM calcium binding which are characterized by domains of 67 amino acids.

Methods

Herein, we accomplished a genome wide comparison and analysis of IQD gene family with in two diploid and tetraploid cotton species. Conserved motifs, gene structure, chromosomal map, evolutionary tree, Ka/Ks distribution, promotor analysis, venn diagram for orthologues genes, and collinearity analysis within genome and between genomes of IQD gene family in four cotton species were performed first time. Expression analysis was performed in four species of cotton from SRA database of NCBI.

Results

293 nonredundant IQD genes were identified from *G. arboreum* (A2), *G. raimondii* (D5), *G. barbadense* (AD2) and *G. hirsutum* (AD1) and clustered in the reference genomes from cotton functional genomics database. 50, 50, 94, and 99 IQD genes were detected from *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* respectively. IQD gene family in four species of cotton elucidated the role of allopolyploid and hybridization during evolutionary cascade of allotetraploid cotton. Comparatively, existence of more orthologous genes in cotton species than Arabidopsis, advocated that polyploidization produced new cotton specific orthologous gene clusters. Duplication of gene events depicts that IQD gene family of cotton evolution was under strong purifying selection. *G. hirsutum* exhibited high level synteny. Expression analysis revealed that GarIQD18, GarIQD25 and GarIQD50 exhibited up regulated expression trend in ovule and fiber in *G. arboreum*. GarIQD25 also exhibited high expression in stem, root and flower. In *G. raimondii*, GraIQD03 demonstrated upregulation expression across stem, ovule, fiber and seed. GbaIQD11 and GbaIQD62 exhibited upregulation fiber development in *G. barbadense*. GhiIQD69 recognized as main candidate genes for plant parts, floral tissues, fiber and ovule development. Promotor analysis identified cis-regulatory elements which are involved in plant defense and stress response mechanisms.

Conclusions

Overwhelmingly, present study paves the way to better understand the evolution of cotton IQD genes and lays a foundation for future investigation of IQD genes in improving abiotic stress tolerance in cotton.

Introduction

Calcium signaling is considered as a messenger that mediates diverse developmental processes and also involved in response to abiotic and biotic stresses (Reddy et al. 2011). Cellular Ca^{2+} signaling utilize their function by spatiotemporal specificity with change in concentration of Ca^{2+} , which could be persuaded by intra-cellular stimuli *i.e.*, pathogenic factors, plant hormones or intra-cellular stimuli like alkali stress, salt, drought and light (Liu et al. 2019; Yuan et al. 2017). Cytoplasmic Ca^{2+} level exhibited ephemeral escalation or fluctuations in regional distribution or their concentration gradient (Yuan et al. 2017). Ca^{2+} receptors (Calcium binding proteins) can decode or detected Ca^{2+} ions, then convert the signals into downstream events which also comprising phosphorylation or dephosphorylation of target proteins (Salveson et al. 2019; Sebastia et al. 2004). Calcium receptors proteins in higher plants could be divided into four groups. Helix loop helix fold motif (EF hand motif) was observed in calcium receptors that behave as calcium binding domains (Wei et al. 2019).

Moreover, EF hand motifs were exhibited by calcium dependent protein kinases (CDPK), calcineurin B like protein (CBL), CaM like protein (CML) and calmodulin (CaM) (Wu et al. 2016). CaM is extensively found in eukaryotic cells and CaM regulated signal transduction pathways responded the main pathway in plants (Wu et al. 2016; Zhang et al. 2014). CaM is well stable and conserved which contain N and C terminal with two spherical domains and connected with a helix motif. EF hand motif is found in every spherical motif that is calcium ion binding domain. Four Ca^{2+} ions can bind with one CaM molecule. Hence changes the CaM conformation to Ca/CaM complex played a key role in target protein and CaM interaction by revealing negatively charged hydrophobic region in spherical surface (Bähler and Rhoads 2002). Apo-calmodulin (ApoCaM) is a CaM without Ca^{2+} ions, transmit signals under low calcium levels by combining calcium independent CaMBP (Feldkamp et al. 2011; Nakanishi et al. 2017). Illustration of analysis and identification of CaMBPs is important to elucidate the detailed signal transduction of CaM and other diverse biochemical and physiological reactions (Zhou et al. 2018). For the modulation of cellular physiology, CaM transmit signals by interrelating with CaM binding proteins because no enzymatic activity demonstrated by CaM (DeFalco et al. 2010). CaMBPs played a very essential role in CaM and Ca^{2+} as well as target proteins. Hence, CaMBPs could be classified into two categories *i.e.*, Ca^{2+} independent and Ca^{2+} dependent (Reddy et al. 2011).

Initially discovered Ca^{2+} independent CaM binding motif was IQ motif. IQ motif containing protein in plants are IQ67 domain containing IQD protein family, IQM protein family, cyclic nucleotide gated channel (CNGC) protein family, calmodulin binding transcription activator (CAMTA) protein family and myosin protein family (Abel et al. 2013). Family members of IQD gene, CML/CaM binding plant specific proteins (CaMBPs) were firstly identified in rice and Arabidopsis (Abel et al. 2005). IQ67 domain is characterized by comprising residues of 67 amino acids. Moreover, IQ motif have very unique and specific repetitive arrangements for example Ca^{2+} dependent CaM recruitment motifs showed 1-8-14 and 1-5-10 arrangements (Bürstenbinder et al. 2013). Previously, IQD gene families were analyzed in *Solanum lycopersicum* (Huang et al. 2013), *Glycine max* (Feng et al. 2014), *Phyllostachys edulis* (Wu et al. 2016), *Oryza sativa* (Abel et al. 2005), *Populus trichocarpa* (Ma et al. 2014), *Brassica rapa ssp. pekinensis* (Duan et al. 2019), *Vitis vinifera* (Liu et al. 2020) and *Arabidopsis*

thaliana (Abel et al. 2005) and functions of limited IQD genes were reported. *SUN* gene in IQD family of tomato reported for organ shape by interrelating with microtubules (Abel et al. 2013). AtIQD5 involved in the process of cell morphogenesis through calcium signaling in *Arabidopsis thaliana* (Liang et al. 2018). SUN24 in IQD gene family in tomato controls germination of seed via abscisic acid signaling pathway (Bi et al. 2018). AtIQD1 gene found in microtubules which interrelate with KLCRQ (kinesin light chain related 1) protein to expedite the cellular transportation (Bürstenbinder et al. 2013). PtiIQD gene family exhibited tissue specific pattern and also found to be involved in methyl jasmonate and drought stresses (Wu et al. 2016). Similarly, GmIQD gene family also expressed specifically and also regulated methyl jasmonate stress (Feng et al. 2014).

Cotton is one of the most important cash crops around the world, because it provides edible oil for food use as well as raw material for textile industry. Hence, production of fiber along with its quality is a prerequisite in cotton. Whereas the production process is significantly affected by biotic and abiotic stresses. Salt stress has been found to reduce fiber production of cotton 20% and aphid infestation also reduce fiber production up to 30% in China (Li et al. 2018; Shaban et al. 2018). IQD family members are also reported to be involved in plant stress signal transduction. Nevertheless, IQD gene family yet not described in cotton species. In present study, initially we identified 293 *Gossypium* IQD genes and evaluated their location of chromosomes, phylogenetic relationships, gene duplication, prediction of physiochemical properties of proteins and expression patterns of floral parts, plant organs, fiber, ovule, and under various stresses. These preliminary findings of *Gossypium* IQD genes pave the way for further research on the biochemical and physiological characterization of IQD protein in cotton.

Materials And Methods

Identification and characterization of cotton IQD genes. The genome assembly sequences of *G. arboreum* (CRI), *G. raimondii* (JGI), *G. barbadense* (ZJU) and *G. hirsutum* (ZJU) were downloaded from cotton functional genomics database <http://cottonfgd.org/>. The targeted candidate genes were explored with known IQD protein sequence of Arabidopsis were taken as reference query by using program BLASTP (Matsuda et al. 2013) with 1e-10 as E value and after that, hidden markov model (HMM) were used to identify all possible IQD sequences with HMMER software (Finn et al. 2011). Moreover, search Pfam tools (<http://pfam.xfam.org/search>) and NCBI CD search (<http://www.ncbi.nlm.nih.gov/cdd/>) were used for the confirmation of candidate genes sequences. Biophysical characteristics of IQD proteins were analyzed with the help of an online tool ExPASy (<http://www.expasy.org/>). The subcellular locations of proteins were calculated using the prediction tool at Bologna Unified Subcellular Component Annotator <http://busca.biocomp.unibo.it>.

Intron/Exon structure and multiple sequence alignment analysis of IQD genes. The IQD genes sequences of 4 species of cotton were aligned with the help of an online tool of Clustal Omega with default settings (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers et al. 2011). Conserved motifs of IQD proteins were identified by MEME version 5.1.1 (Bailey et al. 2015), search was performed with 25 to 200bp for motif width and at least 5 motif sites with 12 motifs count. Identified 12 motifs then searched in protein database in MAST program (Bailey and Gribskov 2000). Coding regions and gnomonic sequences of cotton were assessed to construct gene structure. Furthermore, intron distribution pattern and intron splicing phase were determined by aligned cDNA sequences. Intron/Exon structure of IQD genes was elucidated by tbtools.

Phylogenetic analysis of IQD genes. A total of 326 IQD proteins comprised of 50 IQD genes of *G. arboreum*, 50 of *G. raimondii*, 94 of *G. barbadense*, 99 of *G. hirsutum* and 33 genes of Arabidopsis were used to construct a phylogenetic tree by using MEGA 7.0 by neighbor joining method (Kumar et al. 2016). Bootstrap calculation of tree nodes with 1000 replicates was performed for statistical reliability (Felsenstein 1985). The evolutionary distances were computed using the number of differences method (Nei and Kumar 2000) and are in the units of the number of amino acid differences per sequence.

Chromosomal location, collinearity and multiple synteny analysis of IQD genes. Chromosome locations of IQD genes of four species of cotton were achieved from cottonFGD via genome annotation files. Mapchart was used to visualize and map to each IQD gene on its corresponding chromosome (Caligari and Snape 2004). Gene duplication was assessed through MCScanX (Wang et al. 2012b) and the results were achieved by TBtools (Chen et al. 2018). Gene duplication events were analyzed using the multiple collinearity scan toolkit MCScanX (Wang et al. 2012b). To exhibit segmentally duplicated pairs and orthologous pairs of IQD genes of *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum*, we used multiple synteny plotter software in TBtools (Chen et al. 2018) to draw collinearity maps. The visualize duplicated regions in four species of cotton lines were drawn between matching genes using Circos function built-in TBtools. Substitution rates of synonymous (Ks) and non-synonymous (Ka) values of IQD genes were calculated also by TBtools. The duplication events time (T) was calculated as $T = Ks / 2\lambda \times 10^{-6} (\text{Mya})$, supposing clock-like rate (λ) = 1.5×10^{-8} in cotton (Wang et al. 2012a).

Sequence based orthologous IQD genes identification. The orthologous genes in *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* were characterized using orthoVenn2 (Xu et al. 2019a) (<https://orthovenn2.bioinfotoolkits.net/cluster-venn>). Protein sequences of identified IQDs genes from all four cotton species were used in analysis. Similar calculation was also performed by adding the protein sequences of *A. thaliana*. Moreover, for identification of orthologous genes cluster, each cotton species was individually assessed with each other as well as with Arabidopsis.

Expression profiles analysis in different tissues of cotton IQD genes. Original expression data of *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* IQD genes of various tissues and under heat, cold, PEG (Polyethyl glycol), salinity and control conditions for various hours were retrieved from NCBI database PRJNA594268 for *G. arboreum*, PRJNA490626 *G. barbadense*, PRJNA594268 for *G. hirsutum* and PRJNA171262 for *G. raimondii* from the sequence read archive (SRA) (<https://www.ncbi.nlm.nih.gov>). Adaptors were removed by Trimomatic software (Dong et al. 2016) to perform quality control. Genome reads was mapped by Hisat2 program (Xu et al. 2018), then FPKM values was obtained by using Cufflinks (Jiang et al. 2018). Results were log transformed and Genesis software (Sturm et al. 2002) was used to generate heat map.

Identification Of Cis Regulatory Elements

Upstream sequences of 2000bp of each IQD gene in *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* was retrieved from cottonfgd database <https://cottonfgd.org/analyze/> to assess promotor region. Same upstream sequences were used to trace cis-regulatory elements by using plant care tool <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/> (Lescot et al. 2002).

Results

Identification and sequence analysis of four species of cotton IQD.

Four cotton genomes of two diploid and two tetraploid species were downloaded from cotton functional genomics database and scanned through HAMMER and BLASTP software to identify genes that encode proteins that IQD domain. In present studies, we found 309 IQD proteins initially from four cotton species, after alignment and domain analysis, reduced to 293. 50, 50, 94, and 99 IQD proteins were obtained from *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* genomes respectively. Candidate IQD genes were named according to their chromosome position (Lohani et al. 2019). Amino acid residue length of these cotton IQD proteins ranged from 138 ~ 900 in *G. arboreum*, 133 ~ 901 in *G. raimondii*, 133 ~ 900 in *G. barbadense* and 120 ~ 900 in *G. hirsutum*. Isoelectric points were varied from 5.015 ~ 12.235 in *G. arboreum*, 5.051 ~ 12.167 in *G. raimondii*, 5.050 ~ 12.168 in *G. barbadense* and 5.051 ~ 12.307 in *G. hirsutum*. Correspondingly, molecular weights were distributed from 15.832 ~ 99.234 kDa in *G. arboreum*, 15.392 ~ 99.364 kDa in *G. raimondii*, 15.440 ~ 99.353 kDa in *G. barbadense* and 13.69 ~ 99.361 kDa in *G. hirsutum*. The BUSCA (Bologna Unified Subcellular Component Annotator) website analyzes the IQD protein sequences of cotton species and predicts that most of the genes were reside in the nucleus (Table 1, 2, 3, 4). Subcellular localization revealed that maximum genes resided in nucleus. In *G. arboreum* out of fifty genes, nine genes namely GarIQD04, GarIQD16, GarIQD25, GarIQD34, GarIQD36, GarIQD38, GarIQD41, GarIQD48 and GarIQD49 found in chloroplast. Similarly, in *G. raimondii* out of fifty genes ten genes were observed in cytoplasm *i.e.*, GraIQD05, GraIQD08, GraIQD12, GraIQD23, GraIQD33, GraIQD36, GraIQD40, GraIQD43, GraIQD46 and GraIQD49. While in *G. barbadense* from ninety-four genes, twenty GbaIQD genes were found in chloroplast and one gene GbaIQD32 was observed in endomembrane system. Whereas out of ninety-nine genes of *G. hirsutum*, twenty-one GhiIQD genes were detected in chloroplast and one gene GhiIQD50 in plasma membrane. Complete details of nucleic acid, protein, and genomic sequences of 293 IQDs of candidate genes are listed in Supplementary Table S1.

Table 1
GarlQD identified gene in *Gossypium arboreum*.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelectric Point
Ga01G0605.1	GarlQD01	A01	8,011,882	8,016,349	-	4,468	843	5	280	31.737	29.5	10.725
Ga01G1214.1	GarlQD02	A01	24,314,714	24,319,404	-	4,691	2,481	7	826	91.756	3	6.836
Ga01G2532.1	GarlQD03	A01	109,299,330	109,303,470	+	4,141	1,467	4	488	54.03	42	10.594
Ga02G0048.1	GarlQD04	A02	413,832	414,327	+	496	417	2	138	15.832	22	12.235
Ga02G0677.1	GarlQD05	A02	26,686,866	26,688,255	+	1,390	1,158	3	385	43.257	36	10.745
Ga02G1673.1	GarlQD06	A02	97,896,559	97,899,231	+	2,673	1,401	5	466	51.82	33	10.852
Ga03G0103.1	GarlQD07	A03	859,041	861,454	-	2,414	1,623	5	540	60.775	60	11.799
Ga03G1274.1	GarlQD08	A03	70,880,101	70,882,423	+	2,323	1,443	5	480	52.946	29.5	10.668
Ga04G0800.1	GarlQD09	A04	20,479,259	20,481,800	-	2,542	1,356	6	451	50.075	42	11.156
Ga04G1168.1	GarlQD10	A04	70,411,848	70,414,423	-	2,576	1,425	5	474	52.842	46	11.154
Ga04G2047.1	GarlQD11	A04	97,337,070	97,339,224	+	2,155	1,392	5	463	52.356	34	10.785
Ga05G0277.1	GarlQD12	A05	2,472,556	2,473,965	+	1,410	1,236	3	411	45.777	39	10.955
Ga05G0830.1	GarlQD13	A05	7,220,805	7,225,974	+	5,170	1,305	6	434	47.76	29.5	10.732
Ga05G1403.1	GarlQD14	A05	12,439,111	12,440,931	+	1,821	1,212	3	403	44.77	36.5	10.783
Ga05G1446.1	GarlQD15	A05	12,811,702	12,813,404	-	1,703	1,344	4	447	50.609	36	10.735
Ga05G1676.1	GarlQD16	A05	15,008,149	15,010,784	-	2,636	921	5	306	34.953	37	11.143
Ga05G1990.1	GarlQD17	A05	18,148,173	18,152,549	-	4,377	2,703	6	900	99.234	-21	5.015
Ga05G3216.1	GarlQD18	A05	39,484,737	39,490,027	+	5,291	1,281	5	426	47.772	40.5	11.026
Ga06G0060.1	GarlQD19	A06	360,467	363,561	-	3,095	1,995	5	664	72.562	1.5	6.709
Ga06G1947.1	GarlQD20	A06	121,689,738	121,691,471	+	1,734	1,329	5	442	50.122	32.5	10.457
Ga06G2219.1	GarlQD21	A06	127,368,181	127,369,939	+	1,759	1,161	5	386	44.252	51.5	11.57
Ga07G0098.1	GarlQD22	A07	1,151,451	1,152,928	+	1,478	1,245	3	414	45.351	36	10.883
Ga07G0751.1	GarlQD23	A07	8,936,388	8,949,048	-	12,661	2,055	8	684	76.337	37	10.35
Ga08G1130.1	GarlQD24	A08	57,824,195	57,826,979	-	2,785	1,266	4	421	47.401	13.5	9.566
Ga08G1206.1	GarlQD25	A08	70,730,480	70,734,949	+	4,470	1,293	3	430	48.112	35	10.76
Ga08G1227.1	GarlQD26	A08	75,454,395	75,456,306	-	1,912	1,569	5	522	59.051	38	10.434
Ga08G2108.1	GarlQD27	A08	119,460,271	119,463,403	+	3,133	1,446	4	481	53.443	36	10.477
Ga08G2857.1	GarlQD28	A08	128,498,792	128,500,364	+	1,573	1,077	3	358	39.708	36	11.052
Ga08G2909.1	GarlQD29	A08	128,853,655	128,855,750	+	2,096	1,257	5	418	47.076	43.5	11.167
Ga09G1182.1	GarlQD30	A09	66,349,635	66,354,486	-	4,852	1,311	6	436	47.727	27.5	10.684
Ga09G1374.1	GarlQD31	A09	69,841,682	69,843,440	+	1,759	1,230	5	409	46.624	43.5	10.849
Ga09G1728.1	GarlQD32	A09	75,094,989	75,097,555	-	2,567	1,359	5	452	50.115	32.5	10.835
Ga09G2014.1	GarlQD33	A09	78,091,590	78,095,615	+	4,026	1,386	4	461	51.267	39.5	10.859
Ga09G2369.1	GarlQD34	A09	81,216,217	81,217,821	-	1,605	1,215	3	404	45.031	29.5	10.468
Ga09G2416.1	GarlQD35	A09	81,671,613	81,674,874	+	3,262	1,488	6	495	55.461	57	11.798
Ga10G0119.1	GarlQD36	A10	1,249,595	1,251,007	+	1,413	1,155	4	384	43.281	31	10.568
Ga10G1732.1	GarlQD37	A10	93,692,120	93,695,319	+	3,200	1,539	6	512	57.478	55	11.527
Ga10G2763.1	GarlQD38	A10	126,738,071	126,740,286	-	2,216	936	5	311	35.196	46	11.52
Ga11G2998.1	GarlQD39	A11	114,165,622	114,167,779	+	2,158	1,596	5	531	59.552	49.5	11.11

Note: Fifty GarlQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcript Len protein Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelectric Point
Ga11G3774.1	GarIQD40	A11	121,570,177	121,574,193	-	4,017	2,304	6	767	84.671	44	10.513
Ga11G3948.1	GarIQD41	A11	122,883,816	122,885,541	-	1,726	873	3	290	33.471	44	11.799
Ga12G0300.1	GarIQD42	A12	2,514,543	2,517,017	+	2,475	1,356	4	451	49.903	28.5	10.371
Ga12G0386.1	GarIQD43	A12	3,197,592	3,199,505	+	1,914	1,560	5	519	59.1	39	10.393
Ga12G1753.1	GarIQD44	A12	27,668,056	27,669,899	+	1,844	1,503	5	500	55.783	51	11.264
Ga13G0123.1	GarIQD45	A13	1,312,539	1,314,201	+	1,663	1,158	5	385	44.22	42.5	11.004
Ga13G0853.1	GarIQD46	A13	18,558,034	18,562,133	+	4,100	1,650	6	549	60.054	34.5	10.69
Ga13G0857.1	GarIQD47	A13	19,296,222	19,304,287	+	8,066	1,653	6	550	60.648	36	10.641
Ga13G1324.1	GarIQD48	A13	79,853,374	79,854,617	+	1,244	921	3	306	34.535	37.5	11.139
Ga13G2105.1	GarIQD49	A13	115,497,086	115,498,389	-	1,304	1,131	3	376	43.108	41.5	11.106
Ga13G2341.1	GarIQD50	A13	118,696,264	118,697,885	-	1,622	1,140	5	379	42.636	41.5	11.143
Note: Fifty GarIQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcript Len, protein Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.												

Table 2
GralQD identified gene in *Gossypium raimondii*.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelectric Point
Gorai.001G009300.1	GralQD01	D01	893,556	895,584	+	2,029	1,797	3	415	45.524	39	10.95
Gorai.001G022300.1	GralQD02	D01	2,092,949	2,095,570	-	2,622	1,947	5	437	49.258	42	11.16
Gorai.001G072400.1	GralQD03	D01	7,369,588	7,376,607	-	7,020	2,753	8	650	72.178	37	10.45
Gorai.001G248200.1	GralQD04	D01	50,092,825	50,096,249	+	3,425	2,128	6	479	52.791	30	10.72
Gorai.002G064600.1	GralQD05	D02	7,563,697	7,568,721	-	5,025	1,399	6	285	32.248	30	10.82
Gorai.002G124800.1	GralQD06	D02	18,399,070	18,404,385	-	5,316	3,287	6	838	92.866	-0.5	6.481
Gorai.002G249700.1	GralQD07	D02	61,364,118	61,367,299	+	3,182	1,849	6	466	51.771	30	10.75
Gorai.003G005000.1	GralQD08	D03	272,695	274,017	+	1,323	1,244	2	133	15.392	21	12.16
Gorai.003G061500.1	GralQD09	D03	11,006,480	11,008,407	-	1,928	1,402	4	385	43.264	37.5	10.85
Gorai.003G162000.1	GralQD10	D03	43,059,198	43,063,784	-	4,587	1,919	4	489	54.187	45.5	10.68
Gorai.004G114300.1	GralQD11	D04	26,784,964	26,788,629	+	3,666	1,935	5	445	49.405	19.5	9.853
Gorai.004G121000.1	GralQD12	D04	30,046,656	30,051,405	-	4,750	1,916	4	456	50.763	42.5	10.73
Gorai.004G124400.1	GralQD13	D04	32,364,278	32,366,799	-	2,522	2,229	4	540	61.116	38	10.37
Gorai.004G208400.1	GralQD14	D04	54,112,610	54,116,634	+	4,025	2,360	4	479	53.154	32	10.31
Gorai.004G282000.1	GralQD15	D04	61,372,986	61,375,048	+	2,063	1,585	3	359	39.781	36	10.93
Gorai.004G286700.1	GralQD16	D04	61,705,538	61,707,870	+	2,333	1,504	5	395	44.6	41	11.24
Gorai.005G012200.1	GralQD17	D05	867,614	870,577	-	2,964	2,072	6	541	60.738	58	11.66
Gorai.006G116900.1	GralQD18	D06	36,635,019	36,642,689	-	7,671	2,961	8	436	47.551	29	10.80
Gorai.006G134500.1	GralQD19	D06	38,971,307	38,973,799	+	2,493	1,799	6	414	47.109	42	10.82
Gorai.006G168400.1	GralQD20	D06	42,819,626	42,823,868	-	4,243	2,389	6	452	50.208	32	10.80
Gorai.006G231900.1	GralQD21	D06	48,119,064	48,123,858	+	4,795	2,188	6	514	57.263	60	11.83
Gorai.006G241800.1	GralQD22	D06	48,805,268	48,807,502	-	2,235	1,751	3	436	48.769	25.5	10.13
Gorai.007G019100.1	GralQD23	D07	1,426,115	1,428,051	+	1,937	1,196	3	289	33.02	42	11.60
Gorai.007G035800.1	GralQD24	D07	2,450,957	2,456,180	+	5,224	3,036	8	767	84.862	44.5	10.56
Gorai.007G112400.1	GralQD25	D07	8,684,481	8,687,455	-	2,975	2,413	5	531	59.548	50.5	11.15
Gorai.008G130200.1	GralQD26	D08	37,449,187	37,451,720	-	2,534	2,199	5	501	56.275	52.5	11.34
Gorai.008G259600.1	GralQD27	D08	54,099,749	54,102,170	-	2,422	2,060	5	518	58.961	38	10.36
Gorai.008G268500.1	GralQD28	D08	54,706,004	54,709,223	-	3,220	1,915	5	451	49.989	27	10.29
Gorai.009G027900.1	GralQD29	D09	2,121,651	2,123,460	+	1,810	1,636	3	411	45.705	40	10.95
Gorai.009G082100.1	GralQD30	D09	5,961,338	5,968,758	+	7,421	2,242	8	435	47.975	30	10.84
Gorai.009G137800.1	GralQD31	D09	10,414,411	10,418,189	+	3,779	2,491	4	403	44.794	36	10.87
Gorai.009G142300.1	GralQD32	D09	10,784,688	10,786,734	-	2,047	1,688	4	448	50.93	36	10.76
Gorai.009G164700.1	GralQD33	D09	12,619,468	12,623,594	-	4,127	1,823	6	306	34.941	36	11.11
Gorai.009G196200.1	GralQD34	D09	15,108,108	15,113,298	-	5,191	3,079	6	901	99.364	-21	5.051
Gorai.009G314900.1	GralQD35	D09	29,587,141	29,592,827	+	5,687	1,426	5	426	47.583	36.5	10.91
Gorai.009G420200.1	GralQD36	D09	65,488,488	65,491,914	+	3,427	1,685	4	500	55.942	42.5	10.86
Gorai.010G005800.1	GralQD37	D10	300,155	303,738	-	3,584	2,178	4	691	75.332	10	8.25
Gorai.010G192000.1	GralQD38	D10	54,885,778	54,888,242	-	2,465	1,909	6	442	50.091	30.5	10.43
Gorai.010G204200.1	GralQD39	D10	57,013,544	57,015,413	-	1,870	1,261	5	386	44.262	49.5	11.47

Note: Fifty GralQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcript Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelectric Point
Gorai.011G032700.1	GraiQD40	D11	2,430,343	2,433,788	+	3,446	1,782	6	311	35.148	46	11.52
Gorai.011G135300.1	GraiQD41	D11	20,432,093	20,437,539	-	5,447	2,198	7	523	58.658	52	11.39
Gorai.011G167700.1	GraiQD42	D11	33,944,367	33,945,963	+	1,597	1,113	5	370	41.451	25.5	10.20
Gorai.011G295300.1	GraiQD43	D11	62,580,395	62,582,509	+	2,115	1,597	6	398	45.049	28	10.40
Gorai.012G010000.1	GraiQD44	D12	1,104,927	1,107,524	-	2,598	1,848	5	468	52.993	37	10.79
Gorai.012G093900.1	GraiQD45	D12	16,767,098	16,770,502	+	3,405	2,244	5	484	54.326	46.5	11.22
Gorai.013G013200.1	GraiQD46	D13	889,272	892,589	+	3,318	1,616	7	385	44.135	42	10.97
Gorai.013G082700.1	GraiQD47	D13	11,395,232	11,401,166	+	5,935	2,680	8	552	60.487	36	10.70
Gorai.013G082900.1	GraiQD48	D13	11,515,018	11,520,335	+	5,318	2,119	8	552	60.735	35	10.64
Gorai.013G201400.1	GraiQD49	D13	51,234,150	51,235,701	+	1,552	1,379	3	376	43.003	39.5	11.07
Gorai.013G220500.1	GraiQD50	D13	54,048,645	54,050,927	-	2,283	1,827	5	377	42.472	42.5	11.20
Note: Fifty GraIQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcript Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.												

Table 3
GbalQD identified gene in *Gossypium barbadense*.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelec Point
Gb_A01G0542.1	GbalQD01	A01	6,723,077	6,727,386	-	4,310	858	5	285	32.377	29.5	10.78
Gb_A01G1109.1	GbalQD02	A01	21,733,555	21,738,248	-	4,694	2,514	6	837	92.634	1.5	6.668
Gb_A01G2437.1	GbalQD03	A01	114,083,581	114,086,257	+	2,677	1,401	5	466	51.747	32	10.81
Gb_A02G1412.1	GbalQD04	A02	78,421,274	78,422,832	-	1,559	1,158	3	385	43.219	37	10.79
Gb_A02G2037.1	GbalQD05	A02	101,801,198	101,801,917	-	720	558	3	185	21.586	24.5	11.29
Gb_A03G0230.1	GbalQD06	A03	2,821,987	2,826,122	+	4,136	1,467	4	488	53.977	41	10.52
Gb_A04G0991.1	GbalQD07	A04	63,152,724	63,155,266	+	2,543	1,425	5	474	52.847	45	11.18
Gb_A05G0265.1	GbalQD08	A05	2,645,841	2,647,250	+	1,410	1,236	3	411	45.741	37.5	10.91
Gb_A05G0793.1	GbalQD09	A05	7,444,258	7,449,379	+	5,122	1,260	6	419	46.194	28.5	10.73
Gb_A05G1347.1	GbalQD10	A05	12,581,282	12,583,096	+	1,815	1,209	3	402	44.643	36.5	10.78
Gb_A05G1386.1	GbalQD11	A05	13,022,974	13,024,663	-	1,690	1,344	4	447	50.674	37	10.77
Gb_A05G1598.1	GbalQD12	A05	15,179,353	15,181,988	-	2,636	921	5	306	34.962	37.5	11.14
Gb_A05G1904.1	GbalQD13	A05	18,207,015	18,211,411	-	4,397	2,703	6	900	99.245	-20	5.05
Gb_A05G3064.1	GbalQD14	A05	38,482,914	38,492,664	+	9,751	1,284	5	427	47.694	39	10.96
Gb_A05G3686.1	GbalQD15	A05	82,661,374	82,663,963	+	2,590	1,443	5	480	53.729	43.5	10.97
Gb_A05G4371.1	GbalQD16	A05	106,449,712	106,451,866	+	2,155	1,395	5	464	52.466	35	10.83
Gb_A06G0053.1	GbalQD17	A06	380,402	383,537	-	3,136	1,980	5	659	71.981	1.5	6.709
Gb_A06G1871.1	GbalQD18	A06	110,490,602	110,492,336	-	1,735	1,329	5	442	50.15	33.5	10.49
Gb_A06G1978.1	GbalQD19	A06	112,630,334	112,632,089	+	1,756	1,161	5	386	44.254	51.5	11.57
Gb_A07G0084.1	GbalQD20	A07	1,070,998	1,072,426	+	1,429	1,197	3	398	43.369	35	10.92
Gb_A07G0699.1	GbalQD21	A07	8,650,837	8,655,900	-	5,064	1,956	6	651	72.638	38.5	10.47
Gb_A08G2158.1	GbalQD22	A08	111,612,733	111,615,373	+	2,641	954	4	317	35.328	23.5	10.64
Gb_A08G2877.1	GbalQD23	A08	119,846,913	119,848,493	+	1,581	1,083	3	360	39.785	36.5	11.05
Gb_A08G2928.1	GbalQD24	A08	120,242,566	120,244,662	+	2,097	1,257	5	418	47.015	44.5	11.21
Gb_A09G1299.1	GbalQD25	A09	61,705,888	61,710,739	-	4,852	1,311	6	436	47.763	29	10.73
Gb_A09G1482.1	GbalQD26	A09	64,978,575	64,980,332	+	1,758	1,230	5	409	46.58	43.5	10.85
Gb_A09G1809.1	GbalQD27	A09	69,854,464	69,857,032	-	2,569	1,359	5	452	50.177	34.5	10.86
Gb_A09G2064.1	GbalQD28	A09	72,698,133	72,702,159	+	4,027	1,395	4	464	51.512	41.5	10.92
Gb_A09G2405.1	GbalQD29	A09	75,844,474	75,847,753	+	3,280	1,536	5	511	57.028	59.5	11.94
Gb_A09G2501.1	GbalQD30	A09	76,633,340	76,634,944	-	1,605	1,215	3	404	45.065	30	10.52
Gb_A10G0313.1	GbalQD31	A10	2,692,059	2,694,274	+	2,216	936	5	311	35.215	46.5	11.57
Gb_A10G1649.1	GbalQD32	A10	76,604,401	76,608,108	+	3,708	1,716	6	571	64.281	47.5	11.15
Gb_A10G2849.1	GbalQD33	A10	111,630,402	111,631,814	+	1,413	1,155	4	384	43.16	32	10.59
Gb_A11G0178.1	GbalQD34	A11	1,666,335	1,668,058	+	1,724	882	3	293	33.746	44	11.75
Gb_A11G0329.1	GbalQD35	A11	2,931,133	2,935,172	+	4,040	2,304	6	767	84.854	45	10.53
Gb_A11G1059.1	GbalQD36	A11	9,984,263	9,986,420	-	2,158	1,596	5	531	59.599	48.5	11.1
Gb_A12G1395.1	GbalQD37	A12	75,689,388	75,691,237	-	1,850	1,509	5	502	56.079	53	11.32
Gb_A12G2702.1	GbalQD38	A12	98,471,722	98,473,640	-	1,919	1,560	5	519	59.256	39	10.39

Note: Ninety-four GbalQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcr protein Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelec Point
Gb_A12G2789.1	GbalQD39	A12	99,105,862	99,108,336	-	2,475	1,353	4	450	49.762	26.5	10.28
Gb_A13G0119.1	GbalQD40	A13	1,250,404	1,252,066	+	1,663	1,158	5	385	44.221	40.5	10.94
Gb_A13G0802.1	GbalQD41	A13	17,067,169	17,071,263	+	4,095	1,650	6	549	60.09	36.5	10.71
Gb_A13G0804.1	GbalQD42	A13	17,428,113	17,436,185	+	8,073	1,644	6	547	60.185	35.5	10.65
Gb_A13G1031.1	GbalQD43	A13	40,157,233	40,159,558	-	2,326	1,443	5	480	53.068	28	10.59
Gb_A13G1333.1	GbalQD44	A13	73,974,170	73,975,413	+	1,244	921	3	306	34.467	35.5	11.04
Gb_A13G2121.1	GbalQD45	A13	105,338,039	105,339,342	+	1,304	1,131	3	376	43.079	40.5	11.09
Gb_A13G2303.1	GbalQD46	A13	107,688,511	107,690,120	-	1,610	1,140	5	379	42.522	41.5	11.23
Gb_D01G0558.1	GbalQD47	D01	6,263,717	6,268,153	-	4,437	936	4	311	35.329	32	10.72
Gb_D01G1190.1	GbalQD48	D01	17,586,007	17,590,562	-	4,556	2,514	6	837	92.632	-0.5	6.479
Gb_D01G2509.1	GbalQD49	D01	62,214,682	62,217,393	+	2,712	1,401	5	466	51.781	30	10.75
Gb_D02G0111.1	GbalQD50	D02	842,028	844,426	-	2,399	1,632	5	543	61.051	58	11.74
Gb_D03G0044.1	GbalQD51	D03	280,470	280,950	+	481	402	2	133	15.44	22	12.17
Gb_D03G0610.1	GbalQD52	D03	11,024,653	11,026,165	-	1,513	1,158	3	385	43.307	37.5	10.85
Gb_D03G1768.1	GbalQD53	D03	50,812,429	50,816,564	-	4,136	1,470	4	489	54.181	42.5	10.58
Gb_D04G0096.1	GbalQD54	D04	1,081,628	1,083,769	-	2,142	1,392	5	463	52.373	36	10.81
Gb_D04G0897.1	GbalQD55	D04	17,226,864	17,229,440	+	2,577	1,425	5	474	52.954	44	11.18
Gb_D04G1357.1	GbalQD56	D04	43,015,019	43,017,565	+	2,547	1,413	5	470	52.268	40.5	10.96
Gb_D05G0271.1	GbalQD57	D05	2,499,824	2,501,233	+	1,410	1,236	3	411	45.849	39	10.96
Gb_D05G0781.1	GbalQD58	D05	6,593,641	6,598,825	+	5,185	1,308	6	435	48.077	29.5	10.84
Gb_D05G1350.1	GbalQD59	D05	11,408,580	11,410,395	+	1,816	1,212	3	403	44.75	37	10.92
Gb_D05G1394.1	GbalQD60	D05	11,784,944	11,786,647	-	1,704	1,347	4	448	50.758	37	10.78
Gb_D05G1620.1	GbalQD61	D05	13,812,915	13,815,555	-	2,641	921	5	306	34.947	36	11.11
Gb_D05G1930.1	GbalQD62	D05	16,545,161	16,549,554	-	4,394	2,703	6	900	99.353	-18	5.168
Gb_D05G3034.1	GbalQD63	D05	31,454,708	31,460,214	+	5,507	1,284	5	427	47.626	36.5	10.92
Gb_D06G0054.1	GbalQD64	D06	344,042	347,112	-	3,071	1,995	5	664	72.498	5	7.317
Gb_D06G1941.1	GbalQD65	D06	55,626,244	55,627,985	-	1,742	1,329	5	442	50.107	30.5	10.41
Gb_D06G2053.1	GbalQD66	D06	57,571,555	57,573,314	-	1,760	1,161	5	386	44.124	47.5	11.33
Gb_D07G0093.1	GbalQD67	D07	989,354	990,831	+	1,478	1,248	3	415	45.559	40	11.03
Gb_D07G0225.1	GbalQD68	D07	2,248,875	2,250,862	-	1,988	1,314	5	437	49.27	43	11.18
Gb_D07G0705.1	GbalQD69	D07	7,696,967	7,700,155	-	3,189	1,689	5	562	62.444	37	10.59
Gb_D08G1154.1	GbalQD70	D08	30,119,327	30,122,276	+	2,950	1,338	4	445	49.57	21.5	10.01
Gb_D08G1253.1	GbalQD71	D08	34,569,915	34,574,094	-	4,180	1,293	3	430	48.013	42	11
Gb_D08G1266.1	GbalQD72	D08	37,227,148	37,229,063	-	1,916	1,623	4	540	61.025	40.5	10.41
Gb_D08G2143.1	GbalQD73	D08	59,193,312	59,201,632	+	8,321	1,449	5	482	53.479	34	10.34
Gb_D08G2866.1	GbalQD74	D08	66,864,687	66,866,255	+	1,569	1,080	3	359	39.973	39	11.1
Gb_D08G2921.1	GbalQD75	D08	67,200,787	67,202,820	+	2,034	1,188	5	395	44.767	41	11.24
Gb_D09G1146.1	GbalQD76	D09	38,549,196	38,555,138	-	5,943	1,398	8	465	50.818	32.5	10.92
Gb_D09G1320.1	GbalQD77	D09	40,962,416	40,964,197	+	1,782	1,245	5	414	47.1	40	10.7
Gb_D09G1646.1	GbalQD78	D09	45,013,271	45,016,880	-	3,610	1,485	7	494	54.995	29	10.6

Note: Ninety-four GbalQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcr protein Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelec Point
Gb_D09G2252.1	GbalQD79	D09	50,701,495	50,704,790	+	3,296	1,536	5	511	57.031	59	11.83
Gb_D09G2341.1	GbalQD80	D09	51,369,744	51,371,348	-	1,605	1,215	3	404	44.991	29	10.47
Gb_D10G0319.1	GbalQD81	D10	2,632,089	2,634,305	+	2,217	936	5	311	35.12	46	11.49
Gb_D10G1315.1	GbalQD82	D10	20,932,027	20,935,070	-	3,044	1,572	5	523	58.685	52	11.4
Gb_D10G2806.1	GbalQD83	D10	66,646,224	66,647,678	+	1,455	1,197	4	398	45.121	31	10.53
Gb_D11G0333.1	GbalQD84	D11	2,647,148	2,651,096	+	3,949	2,304	6	767	84.841	44	10.54
Gb_D11G1088.1	GbalQD85	D11	9,192,827	9,194,984	-	2,158	1,596	5	531	59.574	50.5	11.18
Gb_D12G1391.1	GbalQD86	D12	40,230,041	40,231,881	-	1,841	1,506	5	501	56.255	53	11.34
Gb_D12G2705.1	GbalQD87	D12	57,668,879	57,670,797	-	1,919	1,557	5	518	58.977	39	10.41
Gb_D12G2795.1	GbalQD88	D12	58,329,951	58,332,426	-	2,476	1,353	4	450	49.858	26.5	10.3
Gb_D13G0115.1	GbalQD89	D13	952,328	954,313	+	1,986	1,158	5	385	44.189	41	10.91
Gb_D13G0761.1	GbalQD90	D13	11,116,500	11,121,126	+	4,627	1,890	8	629	69.646	42	10.68
Gb_D13G0763.1	GbalQD91	D13	11,262,244	11,266,363	+	4,120	1,659	6	552	60.921	36	10.69
Gb_D13G1069.1	GbalQD92	D13	24,002,355	24,004,678	-	2,324	1,443	5	480	52.927	27	10.62
Gb_D13G2053.1	GbalQD93	D13	54,199,407	54,200,710	+	1,304	1,131	3	376	43.07	39.5	11.08
Gb_D13G2250.1	GbalQD94	D13	56,353,449	56,355,075	-	1,627	1,134	5	377	42.469	40	11.19
Note: Ninety-four GbalQD were identified and their principle transcript ID, gene name, class, chromosomes, start positions, end position, strand, length, transcr protein Length, molecular weight, charge, isoelectric point (pI) and Localization were listed.												

Table 4
GhilQD identified gene in *Gossypium hirsutum*.

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelect Point
Gh_A01G0547.1	GhilQD01	A01	6,727,273	6,730,296	-	3,024	858	5	285	32.3	30.5	10.828
Gh_A01G1101.1	GhilQD02	A01	22,280,616	22,285,308	-	4,693	2,517	6	838	92.76	3.5	6.87
Gh_A01G2332.1	GhilQD03	A01	116,584,431	116,587,131	+	2,701	1,401	5	466	51.68	31	10.735
Gh_A02G0100.1	GhilQD04	A02	860,271	862,684	-	2,414	1,623	5	540	60.67	58	11.725
Gh_A02G1393.1	GhilQD05	A02	83,708,542	83,709,929	-	1,388	1,158	3	385	43.21	37	10.788
Gh_A02G2016.1	GhilQD06	A02	107,932,214	107,932,709	-	496	363	3	120	13.69	18.5	12.307
Gh_A03G0236.1	GhilQD07	A03	3,075,911	3,080,049	+	4,139	1,467	4	488	53.98	41	10.521
Gh_A04G0950.1	GhilQD08	A04	68,542,489	68,545,030	+	2,542	1,404	5	467	51.97	43.5	11.155
Gh_A05G0263.1	GhilQD09	A05	2,539,519	2,540,928	+	1,410	1,236	3	411	45.86	41	11.022
Gh_A05G0790.1	GhilQD10	A05	7,115,341	7,120,508	+	5,168	1,305	6	434	47.79	29.5	10.732
Gh_A05G1339.1	GhilQD11	A05	12,230,416	12,232,232	+	1,817	1,212	3	403	44.74	36.5	10.783
Gh_A05G1378.1	GhilQD12	A05	12,661,612	12,663,300	-	1,689	1,344	4	447	50.74	37	10.774
Gh_A05G1577.1	GhilQD13	A05	14,752,743	14,755,378	-	2,636	921	5	306	34.96	37.5	11.143
Gh_A05G1877.1	GhilQD14	A05	17,740,731	17,745,126	-	4,396	2,703	6	900	99.31	-21	5.051
Gh_A05G3022.1	GhilQD15	A05	37,783,812	37,787,417	+	3,606	1,020	4	339	37.75	35.5	11.457
Gh_A05G3607.1	GhilQD16	A05	86,990,246	86,992,817	+	2,572	1,425	5	474	52.86	45.5	11.181
Gh_A05G4277.1	GhilQD17	A05	109,683,454	109,685,608	+	2,155	1,395	5	464	52.51	34	10.785
Gh_A06G0050.1	GhilQD18	A06	366,150	369,227	-	3,078	1,980	5	659	72.05	1	6.656
Gh_A06G1837.1	GhilQD19	A06	118,612,307	118,614,040	-	1,734	1,329	5	442	50.17	32.5	10.468
Gh_A06G1940.1	GhilQD20	A06	120,662,412	120,664,166	+	1,755	1,149	5	382	43.73	48.5	11.47
Gh_A07G0094.1	GhilQD21	A07	1,024,237	1,025,712	+	1,476	1,245	3	414	45.32	36	10.883
Gh_A07G0227.1	GhilQD22	A07	2,262,737	2,264,723	-	1,987	1,314	5	437	49.26	42	11.195
Gh_A07G0720.1	GhilQD23	A07	8,520,915	8,525,962	-	5,048	1,956	6	651	72.6	37.5	10.435
Gh_A08G1119.1	GhilQD24	A08	59,649,954	59,652,882	-	2,929	1,338	4	445	49.68	17.5	9.793
Gh_A08G1336.1	GhilQD25	A08	88,693,536	88,695,448	+	1,913	1,620	4	539	60.95	36.5	10.315
Gh_A08G2051.1	GhilQD26	A08	116,014,010	116,016,649	+	2,640	954	4	317	35.37	23.5	10.643
Gh_A08G2765.1	GhilQD27	A08	124,217,191	124,218,769	+	1,579	1,083	3	360	39.81	36	11.001
Gh_A08G2816.1	GhilQD28	A08	124,564,402	124,566,497	+	2,096	1,257	5	418	47.02	44.5	11.211
Gh_A09G1181.1	GhilQD29	A09	65,667,778	65,672,629	-	4,852	1,311	6	436	47.75	29	10.733
Gh_A09G1366.1	GhilQD30	A09	69,019,231	69,020,989	+	1,759	1,230	5	409	46.56	43.5	10.881
Gh_A09G1685.1	GhilQD31	A09	73,812,623	73,815,191	-	2,569	1,359	5	452	50.18	33.5	10.821
Gh_A09G1949.1	GhilQD32	A09	76,607,216	76,611,242	+	4,027	1,386	4	461	51.18	40.5	10.909
Gh_A09G2303.1	GhilQD33	A09	79,847,222	79,850,501	+	3,280	1,536	5	511	57	59.5	11.936
Gh_A09G2394.1	GhilQD34	A09	80,601,961	80,603,565	-	1,605	1,215	3	404	45.05	29.5	10.468
Gh_A10G0319.1	GhilQD35	A10	2,624,721	2,626,936	+	2,216	936	5	311	35.19	46.5	11.571
Gh_A10G1560.1	GhilQD36	A10	80,349,259	80,352,483	+	3,225	1,572	5	523	58.68	50	11.348
Gh_A10G2676.1	GhilQD37	A10	114,998,429	114,999,840	+	1,412	1,155	4	384	43.19	32	10.592
Gh_A11G0174.1	GhilQD38	A11	1,547,258	1,548,982	+	1,725	882	3	293	33.83	44	11.804
Gh_A11G0321.1	GhilQD39	A11	2,810,254	2,814,293	+	4,040	2,304	6	767	84.83	42.5	10.499
Gh_A11G1049.1	GhilQD40	A11	9,858,160	9,860,317	-	2,158	1,596	5	531	59.6	48.5	11.096

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelect Point
Gh_A12G1325.1	GhilQD41	A12	82,428,867	82,430,710	-	1,844	1,503	5	500	55.88	53	11.317
Gh_A12G2604.1	GhilQD42	A12	104,369,421	104,371,339	-	1,919	1,560	5	519	59.13	39	10.393
Gh_A12G2690.1	GhilQD43	A12	105,044,849	105,047,320	-	2,472	1,356	4	451	49.88	26.5	10.281
Gh_A13G0121.1	GhilQD44	A13	1,337,621	1,339,281	+	1,661	1,158	5	385	44.22	42.5	11.004
Gh_A13G0768.1	GhilQD45	A13	17,117,354	17,121,448	+	4,095	1,650	6	549	60.06	36.5	10.713
Gh_A13G0771.1	GhilQD46	A13	17,504,018	17,512,100	+	8,083	1,653	6	550	60.67	37.5	10.716
Gh_A13G0956.1	GhilQD47	A13	39,626,224	39,628,546	-	2,323	1,443	5	480	53.07	29	10.668
Gh_A13G1240.1	GhilQD48	A13	71,736,499	71,737,742	+	1,244	921	3	306	34.55	36.5	11.089
Gh_A13G1987.1	GhilQD49	A13	103,040,364	103,041,625	+	1,262	1,089	3	362	41.51	38	11.091
Gh_A13G2167.1	GhilQD50	A13	105,491,347	105,492,807	-	1,461	981	5	326	38.13	43	11.056
Gh_D01G0541.1	GhilQD51	D01	6,095,865	6,099,032	-	3,168	858	5	285	32.29	29.5	10.777
Gh_D01G1150.1	GhilQD52	D01	17,393,990	17,398,544	-	4,555	2,517	6	838	92.82	-0.5	6.479
Gh_D01G2410.1	GhilQD53	D01	63,213,557	63,216,268	+	2,712	1,401	5	466	51.73	31.5	10.774
Gh_D02G0104.1	GhilQD54	D02	852,083	854,480	-	2,398	1,632	5	543	61.08	58	11.695
Gh_D03G0046.1	GhilQD55	D03	288,346	288,826	+	481	402	2	133	15.44	22	12.168
Gh_D03G0607.1	GhilQD56	D03	10,735,371	10,736,757	-	1,387	1,158	3	385	43.31	37.5	10.852
Gh_D03G1733.1	GhilQD57	D03	50,871,724	50,875,853	-	4,130	1,470	4	489	54.21	42.5	10.589
Gh_D04G0099.1	GhilQD58	D04	1,241,060	1,243,204	-	2,145	1,395	5	464	52.5	36	10.807
Gh_D04G0861.1	GhilQD59	D04	17,592,447	17,595,023	+	2,577	1,425	5	474	52.95	44	11.181
Gh_D04G1282.1	GhilQD60	D04	42,532,242	42,534,788	+	2,547	1,413	5	470	52.27	40.5	10.958
Gh_D05G0268.1	GhilQD61	D05	2,291,214	2,292,623	+	1,410	1,236	3	411	45.92	40	10.954
Gh_D05G0787.1	GhilQD62	D05	6,367,633	6,372,818	+	5,186	1,308	6	435	48.16	29.5	10.841
Gh_D05G1342.1	GhilQD63	D05	11,184,427	11,186,242	+	1,816	1,212	3	403	44.65	38	11.02
Gh_D05G1385.1	GhilQD64	D05	11,583,859	11,585,561	-	1,703	1,344	4	447	50.64	38	10.783
Gh_D05G1607.1	GhilQD65	D05	13,599,356	13,601,996	-	2,641	921	5	306	34.95	36	11.111
Gh_D05G1915.1	GhilQD66	D05	16,284,147	16,288,541	-	4,395	2,703	6	900	99.36	-19	5.087
Gh_D05G3028.1	GhilQD67	D05	31,205,225	31,209,983	+	4,759	1,284	5	427	47.74	39.5	11.033
Gh_D06G0037.1	GhilQD68	D06	249,790	252,860	-	3,071	1,995	5	664	72.53	5	7.317
Gh_D06G1866.1	GhilQD69	D06	58,050,617	58,052,355	-	1,739	1,329	5	442	50.1	30.5	10.407
Gh_D06G1971.1	GhilQD70	D06	59,999,190	60,000,949	-	1,760	1,161	5	386	44.14	47.5	11.333
Gh_D07G0100.1	GhilQD71	D07	1,001,526	1,003,003	+	1,478	1,248	3	415	45.44	37	10.864
Gh_D07G0234.1	GhilQD72	D07	2,244,408	2,246,395	-	1,988	1,314	5	437	49.27	43	11.179
Gh_D07G0707.1	GhilQD73	D07	7,488,669	7,493,724	-	5,056	1,953	6	650	72.1	38.5	10.474
Gh_D08G1102.1	GhilQD74	D08	31,292,757	31,296,008	+	3,252	1,338	4	445	49.46	17.5	9.793
Gh_D08G1176.1	GhilQD75	D08	35,819,692	35,823,873	-	4,182	1,371	4	456	50.87	43.5	10.775
Gh_D08G1206.1	GhilQD76	D08	38,678,403	38,680,318	-	1,916	1,623	4	540	60.92	37	10.301
Gh_D08G2064.1	GhilQD77	D08	60,684,657	60,694,782	+	10126	1,527	6	508	56.48	35	10.326
Gh_D08G2761.1	GhilQD78	D08	68,219,518	68,221,086	+	1,569	1,080	3	359	39.87	38	11.042
Gh_D08G2809.1	GhilQD79	D08	68,589,341	68,591,375	+	2,035	1,188	5	395	44.79	42	11.253
Gh_D09G1141.1	GhilQD80	D09	37,185,820	37,190,704	-	4,885	1,311	6	436	47.55	29.5	10.827
Gh_D09G1315.1	GhilQD81	D09	39,558,889	39,560,670	+	1,782	1,245	5	414	47.1	42	10.775

Principle Transcript ID	Rename	Chr	Start	End	Strand	Length (bp)	Transcript Length (bp)	Exon	Protein Length (aa)	Molecular Weight (kDa)	Charge	Isoelect Point
Gh_D09G1630.1	GhilQD82	D09	43,447,899	43,450,453	-	2,555	1,341	5	446	49.66	28	10.775
Gh_D09G2241.1	GhilQD83	D09	49,011,691	49,014,986	+	3,296	1,536	5	511	57.05	59	11.83
Gh_D09G2334.1	GhilQD84	D09	49,672,742	49,674,352	-	1,611	1,221	3	406	45.22	30	10.519
Gh_D10G0332.1	GhilQD85	D10	2,629,194	2,631,409	+	2,216	936	5	311	35.02	45	11.439
Gh_D10G1333.1	GhilQD86	D10	20,467,273	20,470,316	-	3,044	1,572	5	523	58.61	51	11.32
Gh_D10G2782.1	GhilQD87	D10	66,754,786	66,756,238	+	1,453	1,197	4	398	45.12	31	10.528
Gh_D11G0175.1	GhilQD88	D11	1,497,725	1,498,979	+	1,255	882	3	293	33.56	43	11.659
Gh_D11G0334.1	GhilQD89	D11	2,654,638	2,658,586	+	3,949	2,304	6	767	84.78	44	10.541
Gh_D11G1078.1	GhilQD90	D11	9,212,620	9,214,777	-	2,158	1,596	5	531	59.59	49.5	11.137
Gh_D12G1345.1	GhilQD91	D12	41,307,208	41,309,047	-	1,840	1,506	5	501	56.14	52	11.342
Gh_D12G2626.1	GhilQD92	D12	58,557,949	58,559,866	-	1,918	1,557	5	518	58.96	39	10.405
Gh_D12G2716.1	GhilQD93	D12	59,246,779	59,249,257	-	2,479	1,356	4	451	49.96	26	10.253
Gh_D13G0120.1	GhilQD94	D13	1,023,935	1,025,592	+	1,658	1,158	5	385	44.19	41	10.911
Gh_D13G0741.1	GhilQD95	D13	11,568,524	11,572,639	+	4,116	1,659	6	552	60.36	36	10.721
Gh_D13G0744.1	GhilQD96	D13	11,682,596	11,686,716	+	4,121	1,659	6	552	60.9	36	10.687
Gh_D13G1022.1	GhilQD97	D13	25,477,867	25,480,191	+	2,325	1,443	5	480	52.89	27	10.623
Gh_D13G1947.1	GhilQD98	D13	57,285,773	57,287,076	+	1,304	1,131	3	376	43.09	39.5	11.076
Gh_D13G2149.1	GhilQD99	D13	59,505,380	59,507,006	-	1,627	1,134	5	377	42.47	40	11.188

Tree diagram, Gene structure, and conserved motifs of cotton IQD gene family

Neighbor joining method was used to construct an evolutionary tree diagram by using MEGA7. The evolutionary history was inferred using the Neighbor-Joining method [1]. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There was a total of 103 positions in the final dataset. Evolutionary analyses were conducted in MEGA7. There were 103, 84, 89 and 81 positions in the final data set of *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* respectively. All positions gaps and missing data were eliminated with 95% deletion. Results revealed that number of exons varied from 2 ~ 8 in *G. arboreum*, 2 ~ 8 in *G. raimondii*, 2 ~ 8 in *G. barbadense* and 2 ~ 6 in *G. hirsutum* in IQD gene family (Fig. 1, 2, 3, 4). The longest gene length was observed in GarIQD23 with 12,661bp followed by 10,126 bp with GhilQD77bp, 9,751bp in GbalQD14, and 7,671bp in GraIQD18 (Fig. 1, 2, 3, 4). The exons and introns arrangement revealed the evolutionary relationships between various gene family members. Interestingly, closely related genes are structurally more similar just vary in length of intron and exons. Positive association was exhibited by phylogeny and exon-intron structure. Moreover, all the cotton IQD protein sequences were loaded into MEME to uncover the coding motifs. Among them, motif 6 almost remain conserved in all the proteins across 3 species of cotton except *G. barbadense* (Table S2).

Chromosomal location and phylogenetic analysis

Results revealed that 293 cotton IQD genes were unevenly and widely allocated on diploid and allotetraploid chromosomes (Fig. 5a, 5b, 5c). Interestingly, none of the IQD gene was observed on scaffold across four species of cotton on A and D genome (Table 1-4). Number of IQD genes on each chromosome of each species extensively different. The highest number of genes were found on A05 in *G. arboreum*. While highest number of IQD genes was found in *G. raimondii* on D09, *G. barbadense* on A05, A13, D05 and *G. hirsutum* on A05, A13, D05. The largest chromosome size was 135.70 Mb (A03), 85.32 Mb (D13), 120.78 Mb (A08) and 126.48 (A06) in *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* respectively (Fig. 5a, 5b, 5c). Phylogenetic tree of full-length protein sequences was constructed from the multiple sequence alignment of 50 GarIQD, 50 GraIQD, 94 GbalIQD, 99 GhilIQD and 33 AthIQD (Fig. 6). Comparatively, group 6 has the most members of IQD genes followed by group 5, group 1, group 4 and group 3 whilst group 2 contain least members of IQD gene family.

Collinearity and gene duplication of GarIQD, GraIQD, GbalIQD and GhilIQD

Gene duplication always plays a key role in gene expansion and occurrence of novel function of genes. In this study, homologous blast of amino acids of four cotton species was performed with MCScan toolkit. 15891 collinear genes, 1022 collinear blocks and 2375 tandem repeat genes were identified on *G. arboreum* (Table S3). Similarly, 12949, 707 and 1897 collinear genes, collinear blocks and tandem repeat genes were found in *G. raimondii* respectively (Table S3). In *G. barbadense*, 56139 collinear genes, 2215 collinear blocks and 3304 tandem repeat genes were recognized (Table S3). Whereas, 55750 collinear genes, 2191 collinear blocks and 3291 tandem repeat genes were observed on *G. hirsutum* (Table S3). Expansion pattern of IQD genes was studied circos

analysis was performed in four species of cotton (Fig. 7a-d). Paralogous pairs of IQD gene families were arising from whole segmental duplication or genome duplication. Large number of members might arise from transpositions with minor contributions of tandem repeat genes in four species of cotton (Fig. 7a-d). In *G. arboreum*, 11 dispersed, 36 segmental and one proximal duplication was identified (Table S4). Similarly, 24, 2, and 22 genes are exhibited dispersed, proximal and segmental duplication in *G. raimondii*. In tetraploid cotton species, *G. barbadense* showed 1 dispersed, 2 proximal, and 91 segmental gene duplication whilst, 2 proximal and 99 segmental duplication was observed in *G. hirsutum* (Table S4). There are more collinear blocks and genes of A genome in *G. arboreum* and *G. hirsutum* as compared to *G. barbadense* and *G. raimondii*. While similar trend was also observed in D genome comparison (Table S4). Results of homologous blast exhibited the collinearity between *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* (Fig. 8). *G. arboreum* chromosome A05, A09, A10 and A11 share most collinear genes with *G. barbadense* and *G. hirsutum*. Whereas D04, D09, D10 and D11 of *G. raimondii* contributed maximum genes with *G. barbadense* and *G. hirsutum* (Fig. 8).

According to MCScan analysis, 99 gene pairs duplication were detected between diploid *G. arboreum* with tetraploid *G. barbadense* and 101 gene pairs duplication in *G. arboreum* with *G. hirsutum* respectively. Correspondingly, 161, 163 duplicated gene pairs were detected between *G. raimondii* D genome and tetraploid *G. barbadense* and *G. raimondii* with *G. hirsutum* respectively (Table S5). Results also demonstrated that duplication of genes also observed from A genome of *G. arboreum* to A as well as D chromosomes of tetraploid cotton species. Similar behavior was observed in *G. raimondii* with *G. barbadense* and *G. hirsutum* genomes (Fig. 8). *Ka/Ks* ratio predicts the evolutionary story of a gene region or gene (Hurst 2002). Neutral selection occurred when *Ka/Ks* ratio equals to 1, greater than 1 revealed positive selection whilst less than 1 endorsed purifying selection. In present study, *Ka/Ks* ratio for all the IQDs genes in all the species of cotton was less than 1, illustrating purifying selection, except GbaIQD69 in *G. barbadense* and GhiIQD73 in *G. hirsutum*. 20, 19, 43 and 49 gene duplication pairs were identified in *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* respectively (Fig. 9). In *G. arboreum*, segmental duplication of IQD genes has happened from 4.36~32.48 Mya. Similarly, replication of IQD genes took place between 4.14~35.04 Mya in *G. raimondii*. Moreover, in *G. barbadense* and *G. hirsutum* segmental duplication was occurred from 0.36~4.24 (except one gene pair) and 0.57~4.59 Mya (except two gene pairs) respectively (Table S6). Results also highlighted that A and D genome are the progenitors of existing allotetraploid cotton species.

Orthologous gene clusters identification

The relative assessment was ascertained to detect orthologous IQD gene clusters across the five species *G. arboreum*, *G. raimondii*, *G. barbadense*, *G. hirsutum* and *A. thaliana*. This will assist to evaluate polyploidization events in the evolutionary process of IQD gene family in cotton A and D genomes. Identified orthologous gene clusters and their overlap regions of five species are elucidated in Fig. 10. Maximum clusters were recorded from *G. hirsutum*, followed by *G. barbadense*, *G. arboreum* and *G. raimondii*. Results also revealed that presence of 28 orthologous genes clusters in cotton species indicating that polyploidization has resulted in the evolution of new cotton specific orthologues gene clusters. Orthologues gene clusters were also constructed between each species of cotton as well as with *A. thaliana* individually (Table S7). Furthermore, results also emphasized that number of identified orthologues genes decreases with increase in phylogenetic distances. Comparatively, 220, 218, 175 and 171 number of orthologous genes were identified in *G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii* respectively, whereas 82 orthologous genes were identified in *A. thaliana* (Table. S8). 6 in-paralogs were detected in *G. hirsutum* and *G. barbadense* and 13 in-paralogs were spotted in *A. thaliana* whilst no in-paralog was detected in *G. arboreum* and *G. raimondii*. Interestingly, 4,3, and 1 singleton were noticed *G. hirsutum*, *G. barbadense* and *G. raimondii* respectively.

Expression profiles of IQD genes of Cotton

Transcriptome data of cotton was used to explore the variation of expression in IQD genes across tissues. RNAseq data for leaf, stem, root, flower, fiber, ovule, PEG and salinity of *G. arboreum*, seed, fiber and ovule of *G. raimondii*, root, stem, leaf, torus, epicalyx, sepal, petal, filament, anther, ovule and fiber of *G. barbadense* and *G. hirsutum* were downloaded and evaluated. Results revealed that expression pattern of many IQD genes differed substantially in various tissues (Fig. 11). GarIQD18, GarIQD25 and GarIQD50 exhibited up regulated expression trend in 10DPA, 15DPA and 20 DPA of ovule and fiber in *G. arboreum*. GarIQD25 also exhibited high expression in stem, root and flower (Fig.11a). Interestingly, GarIQD22 unveiled up regulation in fiber whilst down regulation in ovule development. In *G. raimondii*, GraIQD03 demonstrated upregulation expression across stem, 0DPA, 3DPA of ovule, 10DPA and 20DPA DPA of fiber and 10DPA, 20DPA, 30DPA and 20DPA of seed. While GarIQD42 didn't express in any tissue (Fig.11b). GbaIQD11, GbaIQD16 and GbaIQD62 exhibited upregulation in fiber development and GbaIQD23, GbaIQD32, GbaIQD82, GbaIQD83 and GbaIQD86 genes were found upregulated in ovule development. GbaIQD51 and GbaIQD79 elected as key gene plant parts development in *G. barbadense*. GbaIQD44 did not expressed in all observed samples (Fig.11c). GhiIQD69 recognized as main candidate genes for plant parts, floral tissues, fiber and ovule development (Fig.10d). Collectively, IQD genes in *G. arboreum* and *G. hirsutum* demonstrated various expression pattern under diverse environmental stresses hence, advocated that these genes were responsive to stress treatments.

Analysis of cis acting elements of HSF gene family in the promotor region

Cis acting regulatory elements were identified in 2000 bp of 5' upstream region by using online tool (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot et al. 2002). Results revealed that cis acting elements are highly consistent the result of evolutionary analysis and gene structure in *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* (Fig. 12a-d). Interestingly, almost similar trend was observed in four species of cotton in identified cis-regulatory elements. CAAT-box, MYB, TATA-box and unnamed-4 were relatively found in higher numbers in all the studied species. Sequences of CAAT-box and TATA-box include functional elements and vital binding sites. CAAT-box and TATA-box long with MYB are also involved in response of plant stresses and defense mechanisms *i.e.*, Drought, Salicylic acid, light and Absciscic acid.

Discussion

Calcium is considered as the most significant messenger and level of calcium can be engendered by extracellular or intracellular stimuli. Similarly, CaM is also one of the vital calcium signaling receptors than mediate various biochemical and physiological reactions. CaM is activated CaM binding proteins (CaMBPs) to regulate physiology of the cell. IQD proteins are plant specific CML/CaM calcium binding which are characterized by domains of 67 amino acids. Transcription factors are the key regulators for crop improvement as well as assist the plant to activate regulatory mechanisms (Jiang et al. 2018; Xu et al. 2018; Zhang et al. 2019). More IQD genes were possessed by sessile plants as compared to other eukaryotes. In present study, initially we identified 293 *Gossypium* IQD genes and evaluated their location of chromosomes, phylogenetic relationships, gene duplication, prediction of physiochemical properties of proteins and expression patterns of floral parts, plant organs, fiber, ovule, and under various stresses. Previously, different number of IQD genes have been reported in various plants namely, 34 SISUN/IQD genes in *Solanum lycopersicum* (Huang et al. 2013), 67 GmlIQD genes in *Glycine max* (Feng et al. 2014), 29 PelIQD in *Phyllostachys edulis* (Wu et al. 2016), 28 OslIQD in *Oryza sativa* (Abel et al. 2005), 38 PtlIQD in *Populus trichocarpa* (Ma et al. 2014), 35 BrlIQD in *Brassica rapa ssp. pekinensis* (Yuan et al. 2019), 49 VviIQD in *Vitis vinifera* (Liu et al. 2020) and 33 AtlIQD in *Arabidopsis thaliana* (Abel et al. 2005) and functions of limited IQD genes were reported. The number of identified IQD genes in two diploid *G. arboreum*, *G. raimondii* and two tetraploid *G. barbadense* and *G. hirsutum* species of cotton is greater than already reported other plant species. Presence of more than ninety IQD genes in tetraploid species of cotton might be due to the hybridization of *G. raimondii* and *G. arboreum* with succeeding polyploidization 1–2 mya 25. Previously, similar results were reported in cotton (Wang et al. 2014) and peanut (Wang et al. 2017b). Comparative analysis of candidate IQD genes indicated that domain conservation and gene structure have been remained conserved during the cotton evolution. These identified genes are further assigned into six groups. Same group shared by many genes in the phylogenetic tree revealed conservation of IQD genes in evolution. Comparatively, group 6 has the most members of IQD genes followed by group 5, group 1, group 4 and group 3 whilst group 2 contain least members of IQD gene family. Similar results were observed in grapevine (Liu et al. 2020) and in maize (Cai et al. 2016). Subcellular localization predictions revealed that most of the IQD genes in cotton are localized in the nucleus. Similar findings were also depicted in soybean (Feng et al. 2014), whereas some of the genes are also observed in extracellular spaces which also reported that ZmHSF11B2b was found in Cytoplasm (Zhang et al. 2020).

Overall, IQD genes expansion in cotton due to the segmental duplication, and tandem duplication also played a minor role. Whereas IQD duplicated gene pairs subjected to purifying selection consequently played a vital role in controlling genomic diversity in natural populations (Cvijović et al. 2018). Similar purifying selection was reported in Chinese cabbage (Yuan et al. 2019), grapevine (Liu et al. 2020), in moso bamboo (Wu et al. 2016) soybean (Lin et al. 2014) and cotton (Rehman et al. 2020). Moreover, present study found that segmental duplication of IQD genes in *G. barbadense* and *G. hirsutum* occurred from 0.36 ~ 4.24 and 0.57 ~ 4.59 Mya respectively. Research reported that divergence time between A and D genome species is 2 to 13 mya (Li et al. 2015). While transoceanic hybridization of A genome with D genome originated *G. hirsutum* (AD1) and *G. barbadense* (AD2) 1 ~ 1.5 million years ago (Wang et al. 2017a). Likewise, duplication events in wheat genome might assist in extension of HSF gene family (Zhou et al. 2019). It was suggested that in angiosperms, IQDs gene expansion occurred in response to whole genome duplications. Likewise, lineage specific whole genome duplications produced number of variants in plant species (Scharf et al. 2012). In allopolyploid process which resulted from fusions of genomes, after duplication, gene retention and gene loss are assumed to enhance resistance against abiotic stresses (Chao et al. 2013; Maere et al. 2005; Yang et al. 2014) hence, redundant duplicates of cotton IQDs may be involved in the regulation of thermotolerance or abiotic stress tolerance. Orthologues genes also indicated the presence of new subclasses of IQD genes in four species of cotton. Similar findings were also reported in brassica (Lohani et al. 2019). Results revealed that expression pattern of many IQD genes differed substantially in various tissues. GarIQD18, GarIQD25 and GarIQD50 exhibited up regulated expression trend in ovule and fiber in *G. arboreum*. In *G. raimondii*, GraIQD03 demonstrated upregulation expression across stem, ovule, fiber and seed. GbaIQD11, GbaIQD16 and GbaIQD62 exhibited upregulation fiber development and GbaIQD23, GbaIQD32, GbaIQD82, GbaIQD83 and GbaIQD86 genes were found upregulated in ovule development. GbaIQD51 and GbaIQD79 elected as key gene plant parts development in *G. barbadense*. GhilIQD69 recognized as main candidate genes for plant parts, floral tissues, fiber and ovule development. Promotor analysis has been revealed that excess of generic CAAT-box and TATA-box along with MYB cis regulatory acting elements accompanying drought inducibility (Xu et al. 2019b; Zhang et al. 2018). Herein, we accomplished a genome wide comparison and analysis of IQD gene family with in two diploid and tetraploid cotton species. Conserved motifs, gene structure, chromosomal map, Ka/Ks distribution, promotor analysis, Venn diagram for orthologues genes, and collinearity analysis within genome and between genomes of IQD gene family in four cotton species were performed first time. Therefore, this experiment will further expand our insights into functional and evolutionary explication of cotton IQD gene family.

Conclusions

In current study, genome wide bioinformatics analysis of IQD genes of cotton was appraised. transcript length, exons, protein length, molecular weight, charge, isoelectric point (pI) and localization of IQDs vary greatly. We classified and identified all possible cotton IQDs and observed paralogous and orthologous gene through amino acid sequence alignment, the findings were valuable for functional study of cotton IQD. Gene structure display analysis showed that IQD cotton genes possessed 2 to 8 exons whilst nearly half number of genes contain 5 exons. Phylogenetic tree, collinearity analysis conserved motifs, Venn diagram for orthologues genes, expression profiling and promotor analysis were also evaluated. Taken altogether, present study provides and extensive resource for understanding the IQD genes in cotton.

Declarations

Ethics approval and consent to participate.

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Provided as electronically supplementary files.

Code Availability

Not applicable

Conflict of interest

Authors declare that they have no conflict of interest for the publication of the manuscript.

Funding

This work was supported by grants from the National Key Research and Development Program of China (Grant No. 2017YFD0101600).

Authors' contributions

AR wrote the initial draft of the manuscript. ZP, HL, GQ, YJ, ZP made all necessary corrections and carried out final editing of manuscript. SH, AQ and XD proofread the manuscript. Final approval for publication was given by the group leader at institute of cotton research XD.

References

- Abel S, Bürstenbinder K, Müller J (2013) The emerging function of IQD proteins as scaffolds in cellular signaling and trafficking Plant Signaling & Behavior 8:e24369
- Abel S, Savchenko T, Levy M (2005) Genome-wide comparative analysis of the IQD gene families in Arabidopsis thaliana and Oryza sativa BMC evolutionary biology 5:72
- Bähler M, Rhoads A (2002) Calmodulin signaling via the IQ motif FEBS letters 513:107-113
- Bailey TL, Gribskov M (2000) Concerning the accuracy of MAST E-values Bioinformatics 16:488-489
- Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite Nucleic acids research 43:W39-W49
- Bi L, Weng L, Jiang Z, Xiao H (2018) The tomato IQD gene SUN24 regulates seed germination through ABA signaling pathway Planta 248:919-931
- Bürstenbinder K et al. (2013) Arabidopsis calmodulin-binding protein IQ67-domain 1 localizes to microtubules and interacts with kinesin light chain-related protein-1 Journal of Biological Chemistry 288:1871-1882
- Cai R et al. (2016) Genome-wide analysis of the IQD gene family in maize Molecular Genetics and Genomics 291:543-558
- Caligari P, Snape J (2004) Mapping quantitative trait loci for ag leaf senescence as a yield determinant in winter wheat under optimal and drought-stressed environments. Euphytica 135: 255263
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs Sci Hortic 67:3948
- Zabeau
- Chao D-Y, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE (2013) Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis Science 341:658-659
- Chen C, Chen H, He Y, Xia R (2018) TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface BioRxiv:289660
- Cvijović I, Good BH, Desai MM (2018) The effect of strong purifying selection on genetic diversity Genetics 209:1235-1278
- DeFalco TA, Bender KW, Snedden WA (2010) Breaking the code: Ca²⁺ sensors in plant signalling Biochemical Journal 425:27-40
- Dong C, Hu H, Xie J (2016) Genome-wide analysis of the DNA-binding with one zinc finger (Dof) transcription factor family in bananas Genome 59:1085-1100
- Duan S, Liu B, Zhang Y, Li G, Guo X (2019) Genome-wide identification and abiotic stress-responsive pattern of heat shock transcription factor family in Triticum aestivum L BMC genomics 20:257
- Feldkamp MD, Yu L, Shea MA (2011) Structural and energetic determinants of apo calmodulin binding to the IQ motif of the NaV1. 2 voltage-dependent sodium channel Structure 19:733-747
- Felsenstein J (1985) Confidence-Limits on Phylogenies—an approach using the bootstrap. evolution 39: 783–791. doi: 10.2307/2408678 View Article PubMed NCBI Google Scholar
- Feng L, Chen Z, Ma H, Chen X, Li Y, Wang Y, Xiang Y (2014) The IQD gene family in soybean: structure, phylogeny, evolution and expression PloS one 9:e110896

- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching *Nucleic Acids Research* 39:W29-W37 doi:10.1093/nar/gkr367
- Huang Z, Van Houten J, Gonzalez G, Xiao H, van der Knaap E (2013) Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato *Molecular genetics and genomics* 288:111-129
- Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution *TRENDS in Genetics* 9:486-487
- Jiang Y et al. (2018) Medicago AP2-domain transcription factor WRI5a is a master regulator of lipid biosynthesis and transfer during mycorrhizal symbiosis *Molecular plant* 11:1344-1359
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets *Molecular biology and evolution* 33:1870-1874
- Lescot M et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences *Nucleic acids research* 30:325-327
- Li F et al. (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution *Nature biotechnology* 33:524-530
- Li J et al. (2018) Aphid parasitism and parasitoid diversity in cotton fields in Xinjiang, China *PloS one* 13:e0207034
- Liang H, Zhang Y, Martinez P, Rasmussen CG, Xu T, Yang Z (2018) The microtubule-associated protein IQ67 DOMAIN5 modulates microtubule dynamics and pavement cell shape *Plant physiology* 177:1555-1568
- Lin Y, Cheng Y, Jin J, Jin X, Jiang H, Yan H, Cheng B (2014) Genome duplication and gene loss affect the evolution of heat shock transcription factor genes in legumes *PloS one* 9:e102825
- Liu T, Zhuang L, Huang B (2019) Metabolic adjustment and gene expression for root sodium transport and calcium signaling contribute to salt tolerance in *Agrostis* grass species *Plant and Soil* 443:219-232
- Liu Z, Haider MS, Khan N, Fang J (2020) Comprehensive Sequence Analysis of IQD Gene Family and their Expression Profiling in Grapevine (*Vitis vinifera*) *Genes* 11:235
- Lohani N, Golicz AA, Singh MB, Bhalla PL (2019) Genome-wide analysis of the Hsf gene family in Brassica oleracea and a comparative analysis of the Hsf gene family in B. oleracea, B. rapa and B. napus *Functional & integrative genomics* 19:515-531
- Ma H, Feng L, Chen Z, Chen X, Zhao H, Xiang Y (2014) Genome-wide identification and expression analysis of the IQD gene family in *Populus trichocarpa* *Plant Science* 229:96-110
- Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes *Proceedings of the National Academy of Sciences* 102:5454-5459
- Matsuda F, Tsugawa H, Fukusaki E (2013) Method for Assessing the Statistical Significance of Mass Spectral Similarities Using Basic Local Alignment Search Tool *Statistics Analytical Chemistry* 85:8291-8297 doi:10.1021/ac401564v
- Nakanishi A et al. (2017) AMP-activated protein kinase-mediated feedback phosphorylation controls the Ca²⁺/calmodulin (CaM) dependence of Ca²⁺/CaM-dependent protein kinase kinase β *Journal of Biological Chemistry* 292:19804-19813
- Nei M, Kumar S (2000) *Molecular evolution and phylogenetics*. Oxford university press,
- Reddy AS, Ali GS, Celesnik H, Day IS (2011) Coping with stresses: roles of calcium-and calcium/calmodulin-regulated gene expression *The Plant Cell* 23:2010-2032
- Rehman A, Atif RM, Qyyum A, Du X, Hinze L, Azhar MT (2020) Genome-wide identification and characterization of HSP70 gene family in four species of cotton *Genomics*
- Salveson I, Anderson DE, Hell JW, Ames JB (2019) Chemical shift assignments of a calmodulin intermediate with two Ca²⁺ bound in complex with the IQ-motif of voltage-gated Ca²⁺ channels (Ca V 1.2) *Biomolecular NMR assignments* 13:233-237
- Scharf K-D, Berberich T, Ebersberger I, Nover L (2012) The plant heat stress transcription factor (Hsf) family: structure, function and evolution *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1819:104-119
- Sebastià CH, Hardin SC, Clouse SD, Kieber JJ, Huber SC (2004) Identification of a new motif for CDPK phosphorylation in vitro that suggests ACC synthase may be a CDPK substrate *Archives of Biochemistry and Biophysics* 428:81-91
- Shaban M, Ahmed MM, Sun H, Ullah A, Zhu L (2018) Genome-wide identification of lipoxygenase gene family in cotton and functional characterization in response to abiotic stresses *BMC genomics* 19:599

- Sievers F et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega Molecular systems biology 7:539
- Sturm A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data Bioinformatics 18:207-208
- Wang J, Sun N, Deng T, Zhang L, Zuo K (2014) Genome-wide cloning, identification, classification and functional analysis of cotton heat shock transcription factors in cotton (*Gossypium hirsutum*) BMC genomics 15:961
- Wang K et al. (2012a) The draft genome of a diploid cotton *Gossypium raimondii* Nature Genetics 44:1098-1103 doi:10.1038/ng.2371
- Wang M et al. (2017a) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication Nature genetics 49:579-587
- Wang P et al. (2017b) Genome-wide dissection of the heat shock transcription factor family genes in *Arachis* Frontiers in plant science 8:106
- Wang Y et al. (2012b) MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity Nucleic acids research 40:e49-e49
- Wei C-C, Fabry E, Hay E, Lloyd L, Kaufman N, Yang Y-P, Stuehr DJ (2019) Metal binding and conformational studies of the calcium binding domain of NADPH oxidase 5 reveal its similarity and difference to calmodulin Journal of Biomolecular Structure and Dynamics
- Wu M, Li Y, Chen D, Liu H, Zhu D, Xiang Y (2016) Genome-wide identification and expression analysis of the IQD gene family in moso bamboo (*Phyllostachys edulis*) Scientific reports 6:24520
- Xu F-C et al. (2018) Heterogeneous expression of the cotton R2R3-MYB transcription factor GbMYB60 increases salt sensitivity in transgenic *Arabidopsis* Plant Cell, Tissue and Organ Culture (PCTOC) 133:15-25
- Xu L et al. (2019a) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species Nucleic acids research 47:W52-W58
- Xu Z, Wang M, Guo Z, Zhu X, Xia Z (2019b) Identification of a 119-bp promoter of the maize sulfite oxidase gene (*ZmSO*) that confers high-level gene expression and ABA or drought inducibility in transgenic plants International journal of molecular sciences 20:3326
- Yang C et al. (2014) Evolution of physiological responses to salt stress in hexaploid wheat Proceedings of the National Academy of Sciences 111:11882-11887
- Yuan J, Liu T, Yu Z, Li Y, Ren H, Hou X, Li Y (2019) Genome-wide analysis of the Chinese cabbage IQD gene family and the response of BrIQD5 in drought resistance Plant molecular biology 99:603-620
- Yuan P, Jauregui E, Du L, Tanaka K, Poovaiah B (2017) Calcium signatures and signaling events orchestrate plant-microbe interactions Current Opinion in Plant Biology 38:173-183
- Zhang H, Li G, Fu C, Duan S, Hu D, Guo X (2020) Genome-wide identification, transcriptome analysis and alternative splicing events of Hsf family genes in maize Scientific reports 10:1-11
- Zhang H, Zheng J, Su H, Xia K, Jian S, Zhang M (2018) Molecular cloning and functional characterization of the dehydrin (IpDHN) gene from *Ipomoea pes-caprae* Frontiers in plant science 9:1454
- Zhang P, Wang R, Ju Q, Li W, Tran L-SP, Xu J (2019) The R2R3-MYB transcription factor MYB49 regulates cadmium accumulation Plant physiology 180:529-542
- Zhang X, Zhou H, Zang X, Gong L, Sun H, Zhang X (2014) MIPS: A calmodulin-binding protein of *Gracilaria lemaneiformis* under heat shock Marine biotechnology 16:475-483
- Zhou M et al. (2019) Genome-wide identification, phylogenetic and expression analysis of the heat shock transcription factor family in bread wheat (*Triticum aestivum* L.) BMC genomics 20:505
- Zhou YP et al. (2018) *Arabidopsis* IQM4, a novel calmodulin-binding protein, is involved with seed dormancy and germination in *Arabidopsis* Frontiers in plant science 9:721

Figures

Figure 1

(A) Evolutionary relationship, (B) structure of conserved motifs and (C) gene architecture of *G. arboreum* IQD gene family

Figure 3

(A) Evolutionary relationship, (B) structure of conserved motifs and (C) gene architecture of *G. barbadense* IQD gene family

Figure 4

(A) Evolutionary relationship, (B) structure of conserved motifs and (C) gene architecture of *G. hirsutum* IQD gene family

Figure 5

a. Chromosomal location of IQD genes in *G. arboreum* and *G. raimondii* b. Chromosomal location of IQD genes in *G. barbadense* c. Chromosomal location of IQD genes in *G. hirsutum*

Figure 6

Phylogenetic tree of IQD gene family with bootstrap values among *G. arboreum*, *G. raimondii*, *G. barbadense* and *G. hirsutum* and *A. thaliana*. The tree was generated by MEGA 7.0 with 1000 bootstrap replicates using the neighbor-joining method.

Figure 7

Duplicated cotton IQD genes based on the collinearity of all chromosomes in (A) *G. arboreum* (B) *G. raimondii* (C) *G. barbadense* and (D) *G. hirsutum*

Figure 8

Sub genome distribution and syntenic assessment IQD genes in cotton. Grey lines revealed collinear block whilst duplicated gene pairs were highlighted with blue lines.

Figure 9

Ka/Ks analysis for gene duplication in *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. hirsutum*.

Figure 10

Clustering of orthologous genes analysis between (A) *G. arboreum*, *G. raimondii*, *G. barbadense*, *G. hirsutum* and *A. thaliana* (B) size of each list and (C) number of elements shared.

Figure 11

a. Expression pattern of IQD genes in *G. arboreum* under various tissues b. Expression pattern of IQD genes in *G. raimondii* under various tissues c. Expression pattern of IQD genes in *G. barbadense* under various tissues. d. Expression pattern of IQD genes in *G. hirsutum* under various tissues.

Figure 12

a. Cis Regulatory elements of IQD gene family in *G. arboreum* b. Cis Regulatory elements of IQD gene family in *G. raimondii* c. Cis Regulatory elements of IQD gene family in *G. barbadense* d. Cis Regulatory elements of IQD gene family in *G. hirsutum*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1.Listofproteinucleicacidandgenomicsequences.xlsx](#)
- [S2.MotifsequenceandmotifInfo.xlsx](#)
- [S3.Analysisofsinglegenomecollinearity.xlsx](#)
- [S4.DUplicatedIQDgenesincotton.xlsx](#)
- [S5.Analysisoftwogenomescollinearity.xlsx](#)
- [S6.Kaksvaluesoffourspeciesofcotton.xlsx](#)
- [S7.Venndiagram.docx](#)
- [S8.Detailelistoforthologousgenecluster.xlsx](#)